HW1.4: Consider

$$\begin{aligned} \text{minimize} \quad & \frac{1}{N} \sum\nolimits_{i \in [N]} f_i(x_i) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{aligned}$$

for $\boldsymbol{A} \in \mathbb{R}^{p \times N}$ and $\boldsymbol{x} = [x_1, \ldots, x_N]^T$.

Complete

(a) Assume strong-convexity and smoothness on f. How would you solve this problem when $N = 1000$?

(b) What if $N = 10^9$?

(c) Can we use Newton's method for $N = 10^9$? Try efficient method for computing $\nabla^2 f(\boldsymbol{x}_k)$ for $p = 1$ and $b = 1$ (probability simplex constraint). Extend it to $1 \leq p \ll N$.

(d) Now, add twice differentiable $r(\boldsymbol{x})$ to the objective and solve (a)-(c).

* duality - convergence rate GD and newton - discuss cost and rate - di

(a) We can solve with gradient descent method.

(b) We can solve by stochastic gradient descent with mini batches so that iterations are easier but more it. will be needed

c) i) No because we need to calculate preconditioning matrix depending on the Hessian like $P = \begin{bmatrix} \nabla^2 f & A \\ A & \nabla^2 f \end{bmatrix}$ where the size is $N+p \times N+p$, in this case we need to find inverse of it with dimensions very high, so computationally very hard.

ii) Perhaps we can replace x with a dual variable v of size p, so that the inverse matrix calc. for the Hessian will get easier.

f) We can add $r(x) = \lambda \|x\|_2^2$, and try to change the condition number. In this way, we can get faster convergence.